DOCUMENT RESUME

ED 329 266 IR 053 457

AUTHOR Seaman, Scott
TITLE In-House CD-ROM Publishing: A Feasibility Report.
PUB DATE Oct 89
NOTE 11p.
PUB TYPE Information Analyses (070) -- Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Electromechanical Technology; *Electronic Publishing; Feasibility Studies; Information Dissemination; *Optical Data Disks; Publishing Industry; Reprography

ABSTRACT

As CD-ROM equipment and products proliferate, there is increasing interest in self-publishing CD-ROM disks. This paper outlines the steps and equipment involved in producing a CD-ROM premaster in-house. A significant portion of this document focuses on the process and feasibility of transferring existing print materials into CD-ROM format. Production of a compact disk is a three stage process involving premastering, mastering, and replication. Premastering, the most lengthy and expensive stage, involves transferring either printed or electronic text and graphics onto 9-track tape to provide input to the mastering process. Mastering transfers the information from that tape to a single "master" compact disk. Replication involves the creation of several stampers from the master and the production of duplicate disks. To contain costs and gain control over the publishing process, premastering is increasingly being done in-house on microcomputer-based turnkey CD-ROM premastering systems. Sophisticated image scanners are linked to the premastering hardware for input of print documents to the CD-ROM database. This document grew from a project at the National Center for Research in Vocational Education to develop a full-text CD-ROM product. The intent of the product was to index and provide full-text retrieval of the center's vocational education research. (7 references) (Author/MAB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED329266

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

IN-HOUSE CD-ROM PUBLISHING: A FEASIBILITY REPORT

SCOTT SEAMAN
THE OHIO STATE UNIVERSITY
UNIVERSITY LIBRARIES
1858 NEIL AVE.
COLUMBUS, OHIO 43210

OCTOBER, 1989

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Scott Seaman

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

BEST COPY AVAILABLE

IR053457

Abstract

As CD-ROM equipment and products proliferate, there is increasing interest in self-publishing CD-ROM disks. This paper outlines the steps and equipment involved in producing a CD-ROM premaster in-house. A significant portion of this document focuses on the process and feasibility of transfering existing print materials into CD-ROM format.

Production of a compact disk is a three stage process involving premastering, mastering, and replication. Premastering, the most lengthy and expensive stage, involves transferring either printed or electronic text and graphics onto 9-track tape to provide input to the mastering process. Mastering transfers the information from that tape to a single "master" compact disk. Replication involves the creation of several stampers from the master and the production of duplicate discs.

To contain costs and gain control over the publishing process, premastering is increasingly being done in-house on microcomputer-based turnkey CD-ROM premastering systems. Sophisticated image scanners are linked to the premastering hardware for input of print documents to the CD-ROM database.

This document grew from a project at the National Center for Research in Vocational Education to develop a full-text CD-ROM product. The intent of the product was to index and provide full-text retrieval of Center's vocational education research.

CD-ROM first became available in 1985 as an alternative to magnetic media for mass storage of digital data. Since then CD-ROM storage has been used for indexes, catalogs, and full-text applications such as the *Oxford English Dictionary* and the *Federal Register*.

Currently, most CD-ROM products are bibliographic tools. Although capable of storing tremendous volumes of text, only about 15% of current products are full-text retrieval systems.[1] Many current applications are duplications of existing online databases. Partly this is because suitable retrieval software is still in its infancy as is the necessary scanning technology to support large-scale transfer of existing print materials to electronic form.

The advantages most often cited of CD-ROM are the flexible retrieval system and economical storage of massive amounts of information. Anything that can be transferred to digital form—sound, color or black and white pictures, text, or animation—can be stored on CD-ROM. Manufacturers claim that a formatted CD-ROM holds 500 megabytes of data (the equivalent of 200,000 typed pages) in ASCII format. But non-ASCII data (scanned text or graphic images) consumes considerably more space. An image can easily consume a half megabyte of disk space, color graphics much more, and sound and animation can be used only in brief segments. So, in actual application, the storage capacity of CD-ROM can be limited to as few as 1,000 images. Text databases of 10 to 500 megabytes are best suited for the medium and account for the majority of the current online products.

This outlines the steps involved in the transfer of existing print materials into CD-ROM format. The focus is on the equipment involved in the in-house premastering phase. Other sources exist that describe the financing and marketing of CD-ROM products.[2]

---

1 Steve Holder, "The New Gutenbergs," in *The CD-ROM Handbook*, ed. Chris Sherman, (New York: McGraw-Hill, 1988), 51.

2 Robin Williamson, "The Cost of Becoming a CD-ROM Publisher," in *CD-ROM Fundamentals*, ed. Charles Oppenheim, (London: Butterworths, 1988), 80; Robert Campbell and Barrie T. Stern, "Publishing on CD-ROM: Some Financial Principles and Market Considerations," in *CD-ROM: Fundamentals to Applications*, ed. Charles Oppenheim, (London: Butterworths, 1988), 220-235.

CD-ROM Publishing

The production stages of a CD-ROM disk are illustrated in Figure 1. The most significant steps are the premastering, mastering, and replication phases. Premastering is the most time-consuming of the three and the only step that is increasingly performed in-house. It culminates with a 9-track tape, a "premaster"—that is, a fully functional preliminary 9-track tape version of the compact disk. The tape is shipped to a mastering facility, such as North American Philips' New York plant, where error-correction is added and the data is transferred to a glass CD master. After brief testing, molds are fabricated from the master and, in a process similar to pressing phonograph records, individual CD-ROM disks are stamped.

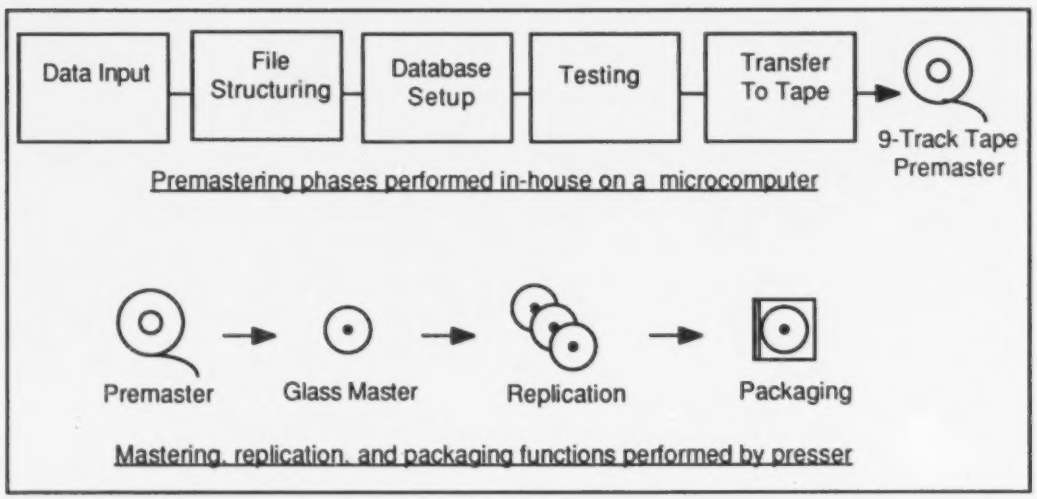

Figure 1. Steps Involved In CD-ROM Publication

In-House Premastering of CD-ROM

Because the premastering process may involve transferring text and graphics from paper to machine-readable form it is the most labor-intensive and expensive phase of CD-ROM production. This stage consists of several steps:

»Data Input
»File Structuring
»Database Setup
»Testing
»Tape Transfer

The first two steps—data input and file structuring—create machine-readable files from paper documents either from scanning or keyboarding and convert them to a common file structure. The database setup procedures add information required by the retrieval software such as index creation (creating the list of locations so data can be found on the disk), data compression (rewriting information into less space), and disk geography creation (physically rearranging data on the disk to speed access of the information). At this point a fully functional version of the database resides on the microcomputer's hard disk and can simulate the functions of the CD-ROM that will be produced. This allows tests of the database in preliminary form for needed changes prior to manufacture. Finally, the tested database is transferred from the microcomputer's hard disk to 9-track computer tape. The resulting premaster is sent to the mastering facility to create the CD master.

In-House CD-ROM Premastering Equipment

Until the past twelve months CD-ROM publishers have depended on outside suppliers, using mini or mainframe configurations for all stages of CD-ROM production. Very recently, however, many CD-ROM publishers have brought the premastering process in-house onto microcomputer-based systems. Since in-house premastering systems utilize modified microcomputers they are significantly less expensive to purchase and operate than mini or mainframe machines and do not require extensive training.

Figure 2 illustrates the the components of an in-house CD-ROM premastering system. The primary components are:

- Microcomputer —IBM PC, Apple Macintosh or comparable machine.
- Hard Disk Units—Approximately 500 megabytes.
- Magnetic Tape Storage Unit—6250 GCR, 1/2 inch, 9-track, 1600 BPI.
- Scanner—Capable of creating ASCII text files and processing graphic images. Requires a separate microcomputer and photocopier.
- Software—Operating system, system software, utilities, and file structure creation programs.

For example, a turnkey CD-ROM publishing system that packages a microcomputer and 9-track tape drive is marketed by Meridian Data, Inc. of Capitola, California. This stand-alone system formats and generates a 9-track premaster. Introduced in September of 1986 the system consists either of a modified IBM PC/AT or an Apple Macintosh microcomputer, a 1600 bits per inch (BPI) 1/2 inch 9-track tape drive and a software package of disk and tape utilities.

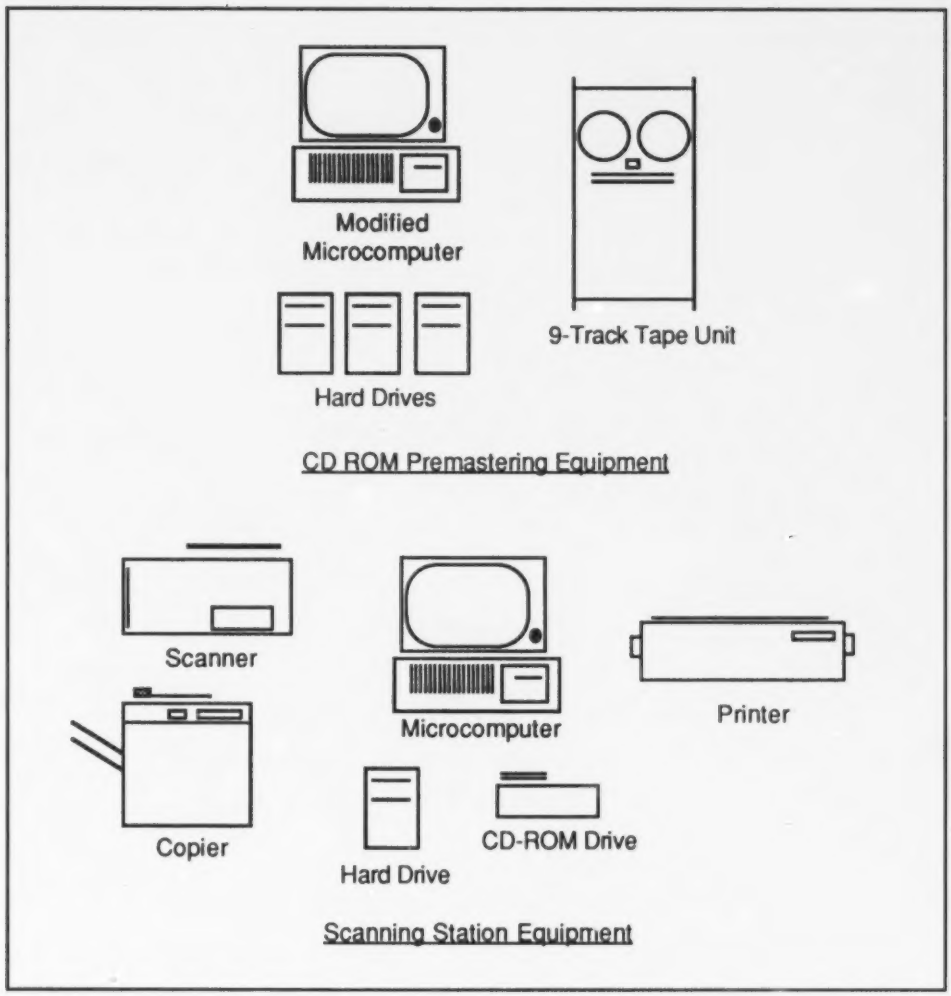

Figure 2. CD ROM Premastering And Scanning Equipment

Microcomputer

The purpose of the microcomputer is to assemble text and graphics files in a format capable of being transferred to a 9-track tape. Either the IBM PS/2 or Apple's Macintosh family is capable of driving the 9-track tape unit. Presently, the IBM is the predominant machine but the Macintosh family, because of superior graphics handling and a more sophisticated operating system, is becoming prevalent. The IBM or Apple must be modified to include 2 megabytes RAM, 120 megabytes of hard drive and, in the case of the IBM, graphics capability.

9-Track Tape Drive

Magnetic tape is the standard data interchange format between large and small computers. Consequently, CD-ROM pressers require premasters be delivered on 9-track 1/2 inch tape recorded at 1600 BPI density.

Scanner Station

The scanner station converts printed text to ASCII format for editing with a word processor and generates machine readable graphics at a resolution of at least 300 dots per inch (DPI) even when text and graphics are mixed on the same page.

Sophisticated scanners are capable of scanning simple pages and multicolumn pages, recognizing common fonts in type sizes ranging from six to twenty-eight points, and can simultaneously scan graphic images at 300 DPI. Typically, a high quality scanner can scan and format a page of text in ASCII within a minute although pages with mixtures of text and graphics take longer. Currently, prices for sophisticated scanners approach $40,000.

A second PC, dedicated to the scanning station, is required for quality control and initial formatting and editing of scanned text and graphics. It must be configured with at least one megabyte RAM, a 20 megabyte hard drive, two SCSI ports and graphics adapter card. A printer is necessary to print scanned text for inspection and a CD-ROM drive would be useful to test CD-ROM masters.

While image scanning has made enormous progress recently, there still exists significant limitations on what can be scanned. Scanners cannot read dot-matrix print or proportionally spaced documents. Soiled, wrinkled, or badly photocopied materials do not scan. Microfiche prints made by wet-chemical process (such as the copies distributed by ERIC) cannot be scanned. Serifed fonts do not scan well, and multiple font styles (italic, boldface, etc.) cannot be reproduced.

Further, scanner output must be carefully examined for missing text, characters that missing or misread (reproducing a "q" rather than "p"), and imperfect graphics. Even sophisticated scanners require quality control checks for every page. Parts of a document may scan without problems while other parts may be unscannable. The publisher is then faced with either manually keying in the unscannable data or leaving it out completely. Finally, scanners are labor intensive because books and reports must be photocopied page-by-page and manually fed into the scanner.

Software

The CD-ROM publishing system will require a variety of software packages. The most significant is the search software to be distributed with each compact disk. Whereas the CD-ROM disk has only database information stored on it, the search

software must be distributed on floppy disk to all users. It contains the search software used to access the database.

Early in CD-ROM development the decision must be made to develop search software or to purchase a software package from an outside company. Owning a proprietary system is potentially a good alternative to buying existing software, but is expensive and time-consuming to develop. Knowledge Access, for example, spent three years creating and testing their CD-ROM retrieval system and must continually develop the software to keep it state of the art.

Consequently, most publishers work with a firm already involved in optical retrieval software because of the cost savings in development, faster delivery time, and the ability to work with proven code. The software firm works with the publisher to design indexes, develop a unique encryption scheme, create menus, match with the software, do initial testing of the software, and provide documentation. This process requires about six months. The cost of rights to full-text retrieval software varies between $10,000 and $50,000 depending on the complexity of the software, the size of the pressing, and the number of editions to be produced.

Conclusions

Having been introduced only a few years ago, CD-ROM technology is still in its infancy. The industry has agreed upon few standards and much of the technology is changing rapidly. Many of the companies supporting CD-ROM publishing (excluding mastering companies such as Sony or Philips) have formed in the past thirty-six months and are marketing unproven equipment and software. Like many developing technologies, CD-ROM publishing is expensive and subject to a great deal of change. Because manufacturers have not been successful in developing industry-wide standards for data formating, search software, or even premastering specifications, the technology has been slower in developing than many anticipated.

While high cost has prevented many cost-sensitive institutions from adopting CD-ROM technology, it is clear that the most limiting factor for CD-ROM publication is the state of scanner technology. Even today's most sophisticated scanners cannot meet the demands that CD-ROM dissemination requires. However, this technology is very close to blossoming and many industry analysts expect functional scanners to appear within the next three to five years.

Yet while CD-ROM technology may require only five more years to stabilize, two developing technologies—Erasable CD and Write Once Read Many (WORM)

disks—promise to eclipse CD-ROM as more flexible and inexpensive methods to disseminate information.

WORM technology and Erasable CD should significantly reduce CD-ROM publishing costs. WORM is essentially CD-ROM that can be written on once, then accessed using a device similar to a CD drive. This technology eliminates the need for a tape premaster and the entire mastering process since disks can be replicated in-house. All other aspects of publication remain the same. WORM disks and equipment are available but are not widely available. Potentially, erasable CD also offers the advantages of CD-ROM in an erasable and reusable format.

CD-ROM, or one of its related technologies, has the potential to alter fundamentally the ways large amounts of information are distributed. However, because of the volatility of compact disk technology, and its expense, investments in CD-ROM publishing should be approached with considerable caution during the next few years.

Selected Bibliography

Andre, Pamela Q. J., and Nancy L. Eaton. 1988. National Agriculture Library text digitizing project. *Library hi tech.* 6: 3, 61-6.

Archer, David. 1987. The challenge to telecommunications: new technologies for large scale local document storage. In *Telecommunications for information management and transfer : Proceedings of the first international conference held at Lancaster Polytechnic, April, 1987.* Brookfield, Vt.: Gower.

Cahan, Mitchell A. 1989. Optical Character Readers/Text Scanners: A Market Analysis. *Information Technology & Libraries.* 8: June 1989, 186-96.

Crossan, David F. 1989. Recognition Technologies. *Inform.* 3: April 1989, 26-8.

Williamson, Robin. 1988. The cost of becoming a CD-ROM publisher. In *CD-ROM: Fundamentals to Applications,* ed. Charles Oppenheim, 80-113. London: Butterworths.

Campbell, Robert, and Barrie T. Stern. 1988. Publishing on CD-ROM: some financial principles and market considerations. In *CD-ROM: Fundamentals to Applications,* ed. Charles Oppenheim, 220-235. London: Butterworths.

Holder, Steve. 1988. The new Gutenbergs. In *The CD-ROM handbook,* ed. Chris Sherman, 51-76. New York: McGraw-Hill.